Introduction to Machine Learning with Python
Fundação Getúlio Vargas

Instructor: Professor Luis Gustavo Nonato

Books:

- Murphy, Kevin P. *Machine learning: a probabilistic perspective*. MIT press, 2012.
- Alpaydin, Ethem. *Introduction to machine learning*. MIT press, 2014.
- Bishop, C. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. *Springer,* 2011.
- Hastie, Tibshirani, and R. Tibshirani. Friedman, J. The Elements of Statistical Learning; Data Mining, Inference and Prediction." 453-480, 2008.
- J. M. Moreira, A. Carvalho, T. Horváth. *A General Introduction to Data Analytics*, *Wiley*, 2019.

Introduction:
welcome, course overview, introducing some basic concepts.
Slides: Intro.pdf

Reading List:
Recommended:
- Murphy's book, chapter 1.
- Alpaydin's book, chapter 1.

Supplementary Material:
- http://scikit-learn.org/stable/

Principal Component Analysis (PCA):
feature extraction and dimensionality reduction
Slides: pca.pdf
Python Codes: Machine_Learning_4_DS-PCA.ipynb

Reading List:
Recommended:
- Murphy's book, section 12.2.
- Bishop's book, sections: 12.1, 12.3

Supplementary Material
- H. Abdi and L.J. Williams. Principal component analysis, Wiley interdisciplinary reviews: computational statistics 2.4 (2010): 433-459.
- J. Shlens. A Tutorial on Principal Component Analysis, *arXiv preprint arXiv:1404.1100,* 2014. (https://arxiv.org/abs/1404.1100)

Regression and Regularization:

Least Square, Ridge Regression and Lasso

Slides: regression.pdf

Python Codes: Machine_Learning_4_DS-Regression.ipynb

Reading List:

Recommended:

- Hastie's book, sections: 3.1 – 3.4

Supplementary Material

- Moreira et al., book, chapter: 8

Clustering and Classification:

Clustering: K-means, Hierarchical

Classification: Naive Bayes classifier, Logistic Regression, SVM

Slides: clustering-classification.pdf

Python Codes: Machine_Learning_4_DS-Clustering-Classification.ipynb

Reading List:

Recommended:

- Bishop's book, sections: 9.1-9.2 (clustering), 7.1, 6.1-6.2 (SVM)
- Murphy's book, sections: 8.1-8.3 (logistic regression), 3.5 (Naïve-Bayes), 11.2 (clustering)

Supplementary Material

- Hastie's book, sections: 13.2 (clustering), 6.6.3 (Naïve-Bayes)
- Alpaydin's book, sections: 7.1-7.4 (clustering), 10.7-10.8 (classification)
- Bishop's book, section: 4.3 (logistic regression),
- Ng, A.Y. & Jordan, M. I. On Discriminative vs. Generative Classifiers: A comparison of Logistic Regression and Naive Bayes, *Neural Information Processing Systems*, 2002.
- Berkhin, P. A survey of clustering data mining techniques. *Grouping multidimensional data*, Springer, 25-71, 2006.
- Hastie's book, sections 12.1-12.3
- Intro to Kernels (Video): https://www.youtube.com/watch?v=kujvLPThqM8

Decision Trees for Prediction and Classification:

Random Forests, Bagging, and Boosting

Slides: decision-trees.pdf

Python Codes: Machine_Learning_4_DS-Decision-Trees.ipynb

Reading List:

Recommended: Hastie's book, section: 9.2, 15.1-15.3, 16.1-16.2

Supplementary Material

- Hastie's book, sections 9.3-9.4, 16.3